Rank Uncertainty in Organizations

Marina Halac Elliot Lipnowski Daniel Rappoport*

December 24, 2019

PRELIMINARY AND INCOMPLETE

Abstract

A principal incentivizes a team of agents to work by privately offering them bonuses contingent on team success. We study the principal's optimal incentive scheme that implements work as a unique equilibrium. This scheme leverages rank uncertainty to address strategic uncertainty. Each agent is informed only of a ranking distribution and his own bonus, the latter making work dominant provided that higher-rank agents work. If agents are symmetric, their bonuses are identical. Thus, discriminatory incentives are strictly suboptimal, in sharp contrast with the case of public contracts (Winter, 2004). We discuss implications for pay transparency and discrimination in organizations.

JEL Codes: D82, D86, L22.

Keywords: mechanism design, information design, contracting with externalities, unique implementation.

---

*Halac: Department of Economics, Yale University, 30 Hillhouse Ave., New Haven, CT 06511, and CEPR (email: marina.halac@yale.edu); Lipnowski: Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027 (email: e.lipnowski@columbia.edu); Rappoport: University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637 (email: daniel.rappoport@chicagobooth.edu). We thank Eyal Winter for helpful comments. We also thank Ale Bonatti and Bob Gibbons for organizing the "Lab on Organizational Economics" at MIT, where we started this project.

1. Introduction

Project success in organizations relies on the contribution of multiple workers whose tasks are complementary. Workers who are rewarded based on overall project outcomes may thus be reluctant to do their share unless they expect that others will as well. We study the optimal provision of incentives that addresses this strategic risk. A firm seeking to ensure effort at the lowest cost must not only offer high enough rewards to make workers' effort worthwhile, but also fine tune rewards to keep workers optimistic that others will also want to work toward their common goal. How does the firm design incentives to manage workers' expectations and facilitate coordination? Is transparency about workers' rank and pay always good? Is pay inequality a feature of optimal incentive provision?

It is intuitive that transparency about coworker effort would help coordination, as workers could then condition their behavior on that of others.[1] Effort however is often not directly observable, neither to the firm nor to other workers, and responding to others' behavior is infeasible when tasks must be undertaken in parallel. Yet, the firm can still provide assurance to a worker that his coworker will work, for example by offering the latter a high enough reward for project success that working is a dominant choice. In fact, based on this logic, Winter's (2004) seminal paper finds that a hierarchy of workers is optimal: high-rank workers are offered steep incentives to work no matter what others do, while low-rank workers are offered just enough to make them work knowing that higher-rank workers will work. Consistent with a broader literature on contracting with externalities (e.g., Segal, 1999, 2003), Winter (2004) obtains that any optimal incentive scheme must be discriminatory, offering differential rewards even when workers are symmetric.

In this paper, we show that the above conclusion hinges on the presumption that contracts within an organization are public, so that workers' relative ranks are transparent. Such a presumption is at odds with evidence and current debates on firm practices. Firms not only rarely disclose information

[1] See Winter (2006, 2010) for theoretical analysis and discussions of the empirical evidence.

about employees' contract terms, but also tend to discourage or even prohibit employees from discussing this information with each other (e.g. Gely and Bierman, 2003; Hegewisch, Williams and Drago, 2011). There is thus often little if any transparency, and this secrecy inside firms is further sustained by social norms (e.g. Edwards, 2005; Cullen and Perez-Truglia, 2018). Taking the canonical team setting of Winter (2004), we show that limiting workers' information about their position relative to others is indeed beneficial to a firm. The firm's optimal incentive scheme creates rank uncertainty in order to better address strategic uncertainty and ensure effort at a lower cost. Furthermore, we prove that this scheme is unique and treats symmetric workers symmetrically. That is, when contract offers can be private, discrimination is not only unnecessary but strictly suboptimal for the firm.

Our model is cast in a moral-hazard-in-teams setting (Holmström, 1982) with effort complementarities. A principal incentivizes a number of agents to work toward a common project. Each agent privately chooses whether to work or shirk, where working is costly and increases the likelihood of project success at a rate that is increasing with other agents' work. The principal can commit to make bonus payments to the agents conditional on the project succeeding. The principal's goal is to uniquely implement work at the least total cost, that is, to specify the least-cost incentive scheme such that all agents working is the unique equilibrium.

We begin by showing that an optimal incentive scheme specifies ranks as in Winter (2004). However, a key difference is that the ranking can be uncertain. Formally, the principal specifies a distribution of type profiles, with each agent observing his realized type only and not those of others (and with any two agents having strictly ordered types). Given his belief over other agents' types, each agent type is then offered the minimum bonus for success that makes him want to work under the assumption that agents of lower type work and the rest do not. Since agents' efforts are strategic complements, an agent who believes to have a lower type relative to others must be offered a higher bonus than an agent who believes to have a higher type. Hence, a relatively lower type corresponds to being assigned a higher rank within the

organization.

Using the structure of ranking schemes, we characterize the principal's optimal value in Theorem 1. The principal's problem can be viewed as choosing an average ranking distribution plus an information structure, the latter determining what each agent learns about the realized ranking from his own type. Theorem 1 shows that providing agents with any private information about the ranking is suboptimal, and therefore the principal's problem reduces to only optimizing over the average ranking distribution. To prove this result, we show that the minimum bonus payment required for an agent to work when only lower types work is convex in the agent's belief about the ranking distribution. Hence, it is optimal to induce the same belief for all types of the same agent.

In Theorem 2, we show that our characterization of the principal's optimal value permits a characterization of optimal incentive schemes. Specifically, we show that there is a unique profile of optimal agent bonuses that solve the principal's problem in Theorem 1, and any optimal sequence of incentive schemes must induce a bonus distribution that converges to a degenerate distribution on this optimal bonus profile. Therefore, optimal incentive contracts are (approximately) unique. Moreover, provided that agents are not too asymmetric, they are faced with uncertainty about their respective ranks, both ex ante and also interim, as agents learn minimal information about others from their own contract offers.

An important corollary of Theorem 2 concerns the case in which agents are symmetric. The principal in this case finds it optimal to induce uniform beliefs over all possible agent rankings, and the unique optimal limiting bonuses are equal across the agents. The intuition turns on how the principal provides assurance to the agents that other agents will work. Making a worker more optimistic that his coworkers will work allows the principal to incentivize his effort at a lower cost; however, this requires that some coworker become more pessimistic, with his incentives then more sensitive to his beliefs about others. Thus, to minimize bonus payments, the principal builds assurance between the workers in a mutual way. The principal's optimal incentive scheme makes

symmetric workers hold the same belief about their rank, and therefore demand the same bonus payment to work.

Our results bring a new perspective on the current debate about firm compensation practices. Concerns about pay discrimination have motivated recent regulation to improve transparency inside firms (e.g. Trotter, Zacur and Stickney, 2017), and have led a number of companies to announce more open internal disclosure policies (Shellenbarger, 2016). Our analysis shows that, from the view of optimal incentive provision, lack of transparency does not generate discriminatory incentives. In fact, together with the results of Winter (2004), we obtain precisely the opposite: a firm's optimal incentive scheme is discriminatory if and only if contracts are required to be public. As a practical implication, this suggests that either measures aimed at improving transparency may be counterproductive at reducing discrimination, or factors others than optimal incentive provision are behind firms' discriminatory practices.

We examine the comparative statics of optimal incentive contracts and discuss a number of extensions of our model. ...

Related literature. Our paper belongs to the literature on contracting with externalities in multi-agent settings, pioneered by Segal (1999, 2003). Most of this literature, including Segal (2003), Winter (2004), Bernstein and Winter (2012), and Halac, Kremer and Winter (forthcoming), studies optimal unique-implementation mechanisms as we do. Except for Winter (2004), however, these papers focus on settings in which agents' choices are observable and bilaterally contractible. Moreover, all of this work requires that contracts be public, thus abstracting from the possibility of information design.[2]

Winter (2004) provides a benchmark for our study by analyzing how to uniquely induce a team of agents to work when only the team's overall success is observable and contractible. Restricting attention to publicly observed

[2] In his section IV, Segal (1999) also considers privately observed contract offers but in a setting in which the principal is unable to commit to a scheme. As Segal (1999) studies equilibria in pure strategies, his agents face no on-path uncertainty about others' offers.

offers, he finds that any optimal incentive scheme must be discriminatory if agents' efforts are strategic complements. This result is related to the optimality of "divide and conquer" strategies in Segal (2003). A main insight from our analysis is that this result is overturned when contract offers are allowed to be private.

Our point that introducing uncertainty can reduce a principal's cost of ensuring effort is related to results in Eliaz and Spiegler (2015) and Moriya and Yamashita (2019). Both of these articles, however, feature different constraints on the space of contracts. In the context of Winter (2004), the analysis in Eliaz and Spiegler (2015) amounts to letting the principal randomize over each agent's bonus offer, but, crucially, with bonus distributions that are required to be independent and identical across the (symmetric) agents. Moriya and Yamashita (2019) extend Winter's (2004) setting by letting effort incentives depend on the realization of a binary state. Taking bonus offers to be fixed and public, they show that the principal's optimal scheme may inform agents asymmetrically about the realized state.

More broadly, our paper is related to recent work studying multi-agent persuasion under adversarial equilibrium selection.[3] Inostroza and Pavan (2018) characterize optimal disclosure policies in a global game, where discriminatory disclosures can be beneficial by a similar "divide and conquer" logic as in the above literature. Mathevet, Perego and Taneva (forthcoming) examine the distributions of agents' beliefs that an information designer can induce, deriving a representation for optimal distributions. Hoshino (2019) shows that agents can be persuaded to take an action profile if this profile satisfies a generalization of risk dominance, while Morris, Oyama and Takahashi (2019) provide a general characterization for the adversarial information design problem in binary-action supermodular games.[4] While our paper differs in many

[3] Our principal's goal of inducing all agents to work as the unique equilibrium is equivalent to inducing all agents to work under adversarial selection, i.e., in the lowest-effort equilibrium of the induced game.

[4] See also Doval and Ely (2019), which characterizes, for given actions and payoffs, all the equilibrium outcomes that are consistent with some choice of information structure and extensive form.

aspects, our main departure from this literature is that we consider the joint design of both transfers and information about said transfers.

Finally, there is related work studying pay transparency and discrimination in organizations from different angles. Cullen and Pakzad-Hurson (2019) analyze the effects of increasing transparency on wage bargaining. Both theoretically and empirically, they find that higher transparency can allow workers to renegotiate to a common wage, while also reducing average wages as firms negotiate more aggressively. Habibi (2019) uses a multidimensional signaling model to study a firm's decision to be transparent about pay, finding that either decision may be optimal depending on the firm's technology.

2. Simple Example

Consider a firm with two workers and a production technology as in Winter (2004). Each worker performs a different task. If a worker works, he completes his task with probability 1; if instead he shirks, the probability that he completes his task is only $p \in (0,1)$. The firm's project succeeds if and only if both workers' tasks are completed. Each worker's cost of working is $c > 0$, and workers make their work decisions simultaneously.

The firm observes only whether the project as a whole succeeds and can commit to make bonus payments to the workers under success. The firm's bonus offers define a game between the workers, where a worker's payoff is equal to his expected bonus payment minus the cost c if he works. The firm's problem is to find the least-cost incentive scheme such that both workers working is the unique equilibrium of the induced game.

Suppose first that the firm commits to public bonus contracts. Note that conditional on the other worker working, a worker is willing to work so long as he is offered a bonus for success no smaller than b_L satisfying $b_L - c = pb_L$, i.e., $b_L = \frac{c}{1-p}$. However, if the other worker shirks, then to work a worker requires a bonus no smaller than b_H satisfying $pb_H - c = p^2 b_H$, i.e., $b_H = \frac{c}{p(1-p)} > b_L$. This implies that while offering bonuses $b_1 = b_2 = b_L$ would suffice to induce an equilibrium in which both workers work, these (or slightly higher)

bonuses would not exclude an equilibrium in which both workers shirk. To uniquely implement work, the firm must make it dominant for one of the workers to work, namely offer a bonus above b_H to one of the workers. A bonus above b_L for the other worker then ensures that he also works, knowing that the first worker always works. It follows that the least-cost public scheme that guarantees work entails a cost (slightly above) $b_H + b_L$ to the firm. As stressed by Winter (2004), this scheme is discriminatory, as workers are treated asymmetrically despite them being identical.

It turns out that the firm can do better by introducing uncertainty. For example, consider a scheme in which worker 1 may be offered a high bonus (slightly above) b_H or a low bonus (slightly above) b_L. Worker 1 is informed of his bonus before choosing whether to work, but worker 2 only knows that worker 1 is offered each bonus with equal probability. Note that if worker 1 is offered b_H, then it is dominant for him to work. Since worker 2 knows that this happens with probability $\frac{1}{2}$, the firm can ensure that worker 2 works by offering him a bonus (slightly above) b_M satisfying

$$\left(\frac{1}{2} + \frac{1}{2}p\right) b_M - c = \left(\frac{1}{2}p + \frac{1}{2}p^2\right) b_M,$$

that is,

$$b_M = \frac{c}{\frac{1}{2}(1-p) + \frac{1}{2}p(1-p)}.$$

Moreover, since worker 2 works under bonus b_M, it is then optimal for worker 1 to also work when he is offered bonus b_L. It follows that the firm guarantees work with this scheme at a cost (slightly above) $\frac{b_H+b_L}{2} + b_M$. Since $b_M < \frac{b_H+b_L}{2}$, the firm's cost is lower than the cost computed above under public offers. Furthermore, since $b_L < b_M < b_H$, the firm's scheme is less discriminatory compared to that above.

Intuitively, given the effort complementarities, the firm wishes to provide assurance to the workers that their coworker will work. A public scheme favors one of the workers in order to guarantee his effort and provide assurance to the other worker. By creating uncertainty, however, the firm can build this

assurance in a mutual way, so that both workers provide assurance to each other. This not only makes the workers' contracts less asymmetric, but also allows the firm to ensure effort at a lower cost. The reason is that a worker's incentive to work increases at a decreasing rate with his belief that the other worker works; hence, assurance is most valuable when a worker is pessimistic about his coworker's effort.

This example serves as a prelude for our analysis in the next sections. Our characterization of the firm's optimal incentive scheme will show more generally how creating rank uncertainty can be beneficial to the firm, and how offering discriminatory incentives becomes detrimental.

3. Model

We study a principal's design of incentives for N agents tasked with working on a common project. With some abuse of notation, let the set of agents be $N = \{1, \dots, N\}$. Each agent $i \in N$ can either work at cost $c_i > 0$ or shirk. Depending on the agents' work decisions, the project can either succeed or fail: the production technology is given by $P : 2^N \to [0, 1]$, where $P(J)$ is the probability that the project succeeds conditional on all agents $i \in J$ working and all others shirking. We make the following assumption on P:

Assumption 1. *P is strictly increasing and strictly supermodular, i.e., for any $J, J' \subseteq N$:*

1. *$P(J) > P(J')$ if $J \supsetneq J'$;*
2. *$P(J \cup J') - P(J) > P(J') - P(J \cap J')$ if J, J' are not nested.*

Part 1 says that effort is productive, that is, the project succeeds with higher probability if more agents choose to work rather than shirk. Part 2 says that efforts are complementary, that is, an agent's marginal product (the marginal effect of his effort on the probability of project success) is larger when more other agents are working.

Agents are protected by limited liability, requiring any payments from the principal to the agents to be nonnegative. Without loss, the principal therefore

only offers the agents success-contingent bonuses, with project failure resulting in no payment to any agent. We allow the principal's bonus offers to be private. Formally, an incentive scheme is $\sigma = \langle q, B \rangle$, where $q \in \Delta(\mathbb{N}^N)$ is a finite-support prior and $B = (B_1, ..., B_N)$ is a bonus rule, with $B_i : \text{Supp}(q_i) \to \mathbb{R}_+$. For convenience, let $T_i^q := \text{Supp}(q_i)$ denote the support of the marginal of q along dimension i. The interpretation is that the principal privately informs each agent $i \in N$ of his *type* $t_i \in T_i^q$ and, through the bonus rule, of his success-contingent bonus $B_i(t_i)$.[5] Hence, agents may face uncertainty about the contracts of other agents; the extent of this uncertainty depends on the principal's choice of q, specifically the correlation between agents' types.

An incentive scheme $\sigma = \langle q, B \rangle$ defines a Bayesian game between the agents. In this game, $\langle (T_i^q)_{i \in N},\ q \rangle$ is a common-prior type space; each agent simultaneously makes a type-contingent decision of whether to work or shirk; and an agent's payoff is equal to his expected bonus payment net of any cost of effort he bears. An equilibrium induced by an incentive scheme is a Bayesian Nash equilibrium of the Bayesian game defined by the scheme. The principal's goal is to offer a least-cost incentive scheme that ensures work, namely that induces all types of all agents to work as the unique equilibrium. For expositional convenience when dealing with indifferent agents, we require schemes to induce such a unique equilibrium only once the bonus offers are increased by any positive amount. More precisely, say that $\sigma = \langle q, B \rangle$ **uniquely implements work (UIW)** if, for every $\varepsilon > 0$, all agent types working is the unique equilibrium of the game induced by scheme $\langle q, B + \varepsilon \rangle$ (in which each type t_i of agent i is offered bonus $B_i(t_i) + \varepsilon$).[6] The principal then solves the following problem:

$$\inf_{\sigma \text{ UIW}} W(\sigma), \tag{1}$$

where $W(\sigma)$ is the principal's total expected cost given incentive scheme σ

[5] Since the type itself is simply a label, it is immaterial that agents' types are labeled with natural numbers: drawing each agent's types from any infinite set would keep the principal's problem unchanged.

[6] As will be clear, our characterization results in Theorem 1 and Theorem 2 would remain true as stated if ε were replaced with 0 in this definition.

that uniquely implements work:

$$W(\sigma) = P(N) \sum_{i \in N} \sum_{t_i \in T_i^q} q_i(t_i) B_i(t_i).$$

It will follow from our results that the principal's problem does not generally admit a minimum. Hence, we will focus on approximately optimal incentive schemes. Let W^* denote the principal's optimal value, i.e., the value of program (1). We define an **optimal sequence of incentive schemes** as a sequence $(\sigma^m)_m$ such that each σ^m uniquely implements work and $W(\sigma^m)$ converges to W^* as $m \to \infty$.

A feature of approximately optimal incentive schemes in which we are particularly interested is the distribution of bonuses offered to each agent. Given an incentive scheme $\sigma = \langle q, B \rangle$, let $\beta^\sigma \in \Delta(\mathbb{R}_+^N)$ be the distribution of bonus profiles, i.e., the finite-support distribution with[7]

$$\beta^\sigma(b) := q\left\{t \in T^q : \ B_1(t_1) = b_1, \ldots, B_N(t_N) = b_N\right\} \text{ for all } b \in \mathbb{R}_+^N.$$

Given a sequence $(\sigma^m)_m$ of incentive schemes, $\beta^* \in \Delta(\mathbb{R}_+^N)$ is the limit bonus distribution of this sequence if it is the limit of the sequence $(\beta^{\sigma^m})_m$.[8] Moreover, β^* is an optimal bonus distribution if it is the limit bonus distribution of an optimal sequence of incentive schemes.

Remark 1. We focus our exposition around the principal inducing every agent to work as the unique equilibrium. In our setting, this requirement is equivalent to inducing every agent to work as the unique rationalizable outcome. This follows from the fact that, given the supermodular technology P and the agents' limited liability, the game played by the agents under any incentive scheme is a supermodular game, and thus the results of Milgrom and Roberts (1990) apply.[9] In particular, the strategy profile in which every agent

[7] Note that using this definition, the principal's total expected cost in (1) can be written as $W(\sigma) = P(N) \sum_{i \in N} \int b_i \, \mathrm{d}\beta_i^\sigma(b_i)$, where $\beta_i^\sigma \in \Delta(\mathbb{R}_+)$ is the marginal distribution over the bonus offered to agent i.

[8] Here, we take convergence in the weak* sense.

[9] Milgrom and Roberts (1990) study games of complete information. To apply their results

type chooses his lowest (i.e., least-work) interim correlated rationalizable action is an equilibrium, and if this lowest rationalizable strategy profile does not consist of everybody working, then it cannot be the unique equilibrium for everybody to work.

4. Optimal Incentives

We characterize optimal incentives by solving the principal's problem in (1). We proceed as follows. First, we show that it is without loss for the principal to focus on a simple class of incentive schemes which we call ranking schemes. Second, using the structure of ranking schemes, we derive an auxiliary optimization program that fully characterizes the principal's value of (1). Finally, we show that this auxiliary program also characterizes the solution to (1), by pinning down the limiting bonus distribution of an optimal sequence of incentive schemes. We discuss the implications of optimal bonuses for pay discrimination and study their comparative statics.

4.1. Ranking Agents

It will be useful for our analysis to define a class of incentive schemes that we call ranking schemes:

Definition 1. *An incentive scheme $\sigma = \langle q, B \rangle$ is a **ranking scheme** if:*

1. *No two agents are assigned the same type. That is, for all $i, j \in N$,*

$$q\{t \in T^q : \ t_i = t_j\} = 0 \text{ if } i \neq j.$$

2. *The bonus for each agent type makes him indifferent between working and shirking given the belief that all other agents of lower type work and the rest shirk.*

to our setting, however, note that our game can be viewed in "agent form" by taking each of the finitely many types of each agent as a distinct player. Interim correlated rationalizable play in the induced Bayesian game is equivalent to correlated rationalizable play in the agent-form game, and the latter is a game of strategic complements.

That is, for all $i \in N$ and all $t_i \in T_i^q$,

$$B_i(t_i) \sum_{t_{-i}} q_i(t_{-i}|t_i) \Big[P\{j \in N : \ t_j \leq t_i\} - P\{j \in N : \ t_j < t_i\} \Big] = c_i.$$

Note that the bonuses in a ranking scheme $\sigma = \langle q, B \rangle$ are pinned down, given the prior q, by the indifference conditions in part 2 of Definition 1. Thus, we will sometimes refer to a ranking scheme simply by a distribution q over type profiles (satisfying part 1 of Definition 1), with the understanding that the associated bonus rule is as specified in this definition.

Ranking schemes have a simple structure. Under any given such scheme $\sigma = \langle q, B \rangle$, each type of every agent is indifferent over working if all (strictly) lower types of other agents work and the rest do not. This means that under $\langle q, B + \varepsilon \rangle$, for any $\varepsilon > 0$, each type finds it dominant to work if all lower types work and the rest do not. Moreover, by supermodularity of P, working is then dominant for each type if *at least* all lower types work. It therefore follows, by induction on the type, that any ranking scheme uniquely implements work: under $\langle q, B + \varepsilon \rangle$, the lowest type finds work dominant; the second lowest type finds work dominant given that the lowest type works; the third lowest type finds work dominant given that the lowest and second lowest types work; and so on.

The next proposition shows that ranking schemes are useful not only because they ensure that all agents work, but because they can also be constructed to ensure work at the least total cost to the principal. Hence, to solve the principal's problem in (1), it is without loss to focus on ranking schemes.

Proposition 1. *Every ranking scheme uniquely implements work. Moreover, if an incentive scheme σ uniquely implements work, then there is a ranking scheme σ^* such that, for each agent $i \in N$, the marginal bonus distribution β_i^σ first-order-stochastically dominates $\beta_i^{\sigma^*}$. Hence, the principal's optimal value satisfies*

$$\inf_{\sigma \text{ UIW}} W(\sigma) = \inf_{\sigma \text{ is a ranking scheme}} W(\sigma).$$

For intuition, fix an incentive scheme $\sigma = \langle q, B \rangle$ that uniquely implements work. We can construct a ranking scheme that (weakly) improves the principal's objective in three steps. First, we iteratively assign each type of each agent an "index": the index-1 types are those who find work weakly dominant; the index-2 types are those who find work weakly dominant given that index-1 types work; the index-3 types are those who find work weakly dominant given that index-1 and index-2 types work; and so on. We observe that *every* type will be assigned such an index, for otherwise one can construct an equilibrium for $\langle q, B + \varepsilon \rangle$, with $\varepsilon > 0$ small enough, in which the unindexed types shirk. Second, we relabel types, so that no two agents ever have the same type and, moreover, $t_i < t_j$ whenever t_i has a strictly lower index than t_j. We can then assign bonuses to relabeled types as required by the definition of a ranking scheme, namely as specified in part 2 of Definition 1. Third, we show that, by supermodularity of P, this adjustment to the bonuses systematically (weakly) lowers each type's bonus payment, thereby stochastically reducing the bonus distribution. It follows that, relative to the original scheme, the resulting ranking scheme reduces the principal's total expected cost.

Agents' incentives to work in a ranking scheme depend on the probability that they assign to other agents having realized a lower type than their own. To make these incentives precise, we introduce the following notation. Let Π denote the set of all permutations on N, i.e., all $\pi \in N^N$ such that $\pi_i \neq \pi_j$ for distinct $i, j \in N$. For a type profile t with $t_i \neq t_j$ for all $i, j \in N$, let the **agent ranking** given t be $\pi(t) \in \Pi$ with $\pi_i(t) = |\{j \in N : \ t_j \leq t_i\}|$. Then given a ranking scheme q, agent $i \in N$, and type $t_i \in T_i^q$, let t_i's **ranking belief** $\mu_i^q(\cdot|t_i) \in \Delta\Pi$ be given by

$$\mu_i^q(\hat{\pi}|t_i) := q_i\left(\{t_{-i} : \ \pi^q(t_i, t_{-i}) = \hat{\pi}\} \,\middle|\, t_i\right) \text{ for all } \hat{\pi} \in \Pi.$$

Finally, let the **ranking distribution** $\mu^q \in \Delta\Pi$ be given by

$$\mu^q(\hat{\pi}) := q\left(\{t : \ \pi^q(t) = \hat{\pi}\}\right) \text{ for all } \hat{\pi} \in \Pi.$$

For any agent $i \in N$ and ranking belief $\mu_i \in \Delta\Pi$ that he might hold, we

can compute explicitly the bonus payment required by this agent to work. Specifically, if agent i believes that the set of other agents who work is $J \subseteq N \setminus \{i\}$ with probability $\mu_i\{\pi \in \Pi : \text{ for all } j \in N,\ \pi_j < \pi_i \text{ if and only if } j \in J\}$, then the success-contingent bonus that keeps him exactly indifferent between working and shirking is equal to:

$$\frac{c_i}{\sum_{\pi \in \Pi} \mu_i(\pi) \left[P\{j \in N : \ \pi_j \leq \pi_i\} - P\{j \in N : \ \pi_j < \pi_i\}\right]}.$$

Moreover, since the principal makes the bonus payments only when the project succeeds, and the probability of project success (given that work is uniquely implemented) is equal to $P(N)$, it follows that the principal's expected bonus payment to agent i, as a function of this agent's ranking belief μ_i, is:

$$f_i(\mu_i) := \frac{c_i P(N)}{\sum_{\pi \in \Pi} \mu_i(\pi) \left[P\{j \in N : \ \pi_j \leq \pi_i\} - P\{j \in N : \ \pi_j < \pi_i\}\right]}. \tag{2}$$

The function $f_i : \Delta\Pi \to \mathbb{R}_{++}$ describes the expected fees associated with providing incentives to agent i for any given ranking belief that he might hold. Note that since P is supermodular, an agent who believes himself to be placed earlier (respectively, later) in the agent ranking on average will necessitate higher (respectively, lower) bonuses to be willing to work. Agents who are placed earlier in the ranking thus correspond to "higher-rank" agents who are paid more.

4.2. Principal's Value

By Proposition 1, it is without loss for the principal to restrict attention to ranking schemes when solving for a least-cost incentive scheme that uniquely implements work. Moreover, we have shown that given a ranking scheme q, the bonus offer for type $t_i \in T_i^q$ of agent $i \in N$ can be computed as $\frac{1}{P(N)} f_i\left(\mu_i^q(\cdot|t_i)\right)$, where the function f_i is defined in (2) and $\mu_i^q(\cdot|t_i)$ is t_i's ranking belief. Hence, the principal's total expected cost under ranking scheme q can be written as

$$W(q) = \sum_{t \in T^q,\ i \in N} q(t) f_i\left(\mu_i^q(\cdot|t_i)\right).$$

It is clear from this expression that the relevant choice for the principal is a profile of distributions over ranking beliefs. However, the principal is constrained: she cannot make an agent believe that he is placed later in the agent ranking (thereby reducing his associated bonus) without making some other agent believe that he is placed earlier. To incorporate this constraint, consider the following interpretation of the principal's problem. The principal first chooses an average ranking distribution and then chooses an information structure for each agent about the realized ranking—with the agent's type corresponding to his signal. Every ranking scheme naturally implies some selection of ranking distribution and information structure, and it turns out that a certain converse also holds. Specifically, the following theorem shows that (i) providing no private information about the ranking to any agent is optimal given a ranking distribution, and (ii) such an informational choice imposes no constraints on what the ranking distribution can be. Therefore, the principal's problem can be recast as an unconstrained optimization over ranking distributions themselves.

Theorem 1. *The principal's optimal value satisfies*

$$\inf_{\sigma \text{ UIW}} W(\sigma) = \min_{\mu \in \Delta\Pi} \sum_{i \in N} f_i(\mu).$$

Proof. Note that a minimizer to $\sum_{i \in N} f_i$ exists and each f_i is bounded because the function is continuous over its compact domain.

We first show that $\inf_{\sigma \text{ UIW}} W(\sigma) \geq \min_{\mu \in \Delta\Pi} \sum_{i \in N} f_i(\mu)$. Given Proposition 1, it is sufficient to show that the principal's value for a ranking scheme q is no less than $\min_{\mu \in \Delta\Pi} \sum_{i \in N} f_i(\mu)$. Bayesian updating implies that a given agent i's ranking belief is, on average, equal to the true ranking distribution:

$$\sum_{t \in T^q} q(t) \mu_i^q(\cdot|t_i) = \sum_{t_i \in T_i^q} q_i(t_i) \mu_i^q(\cdot|t_i) = \mu^q.$$

Hence, since f_i is convex, Jensen's inequality implies

$$\sum_{t \in T^q} q(t) f_i(\mu_i^q(\cdot|t_i)) \geq f_i\left(\mu^q\right).$$

Summing over $i \in N$ yields $W(q) \geq \sum_{i \in N} f_i(\mu^q)$, proving the claim.

We next show that $\inf_{\sigma \text{ UIW}} W(\sigma) \leq \min_{\mu \in \Delta\Pi} \sum_{i \in N} f_i(\mu)$. Fix $\mu \in \Delta\Pi$. We construct a sequence $(q^m)_m$ of ranking schemes such that $W(q^m) \to \sum_{i \in N} f_i(\mu)$ as $m \to \infty$. For each $m \in \mathbb{N}$, define the finite-support prior $q^m \in \Delta(\mathbb{N}^N)$ as follows

$$q^m(t) := \begin{cases} \frac{1}{m}\mu(\pi) & : \exists \ell \in \{0, \ldots, m-1\} \text{ with } t_i = \ell + \pi_i \text{ for all } i \in N, \\ 0 & : \text{otherwise.} \end{cases}$$

By construction, for each $m \in \mathbb{N}$ and $i \in N$, every type $t_i \in T_i^{q^m}$ with $N \leq t_i \leq m$ has ranking belief $\mu_i^{q^m}(\cdot|t_i) = \mu$ and thus is offered bonus $\frac{1}{P(N)} f_i(\mu)$. Note that such types of agent i arise under q^m with probability $\max\left\{0, \frac{m-N+1}{m+N-1}\right\}$, which converges to 1 as $m \to \infty$. Hence, the expected bonus payment to each agent i converges to $f_i(\mu)$, proving the claim. *Q.E.D.*

Theorem 1 reduces the principal's problem to choosing (without restriction) a ranking distribution, as opposed to choosing (with some restrictions) a profile of distributions over ranking distributions. This means that it is optimal for the principal not to provide any private information to the agents about the realized ranking. The reason lies in the fact that the minimum bonus payment required for an agent to work is convex in the agent's belief about the ranking distribution. Hence, it is optimal to induce the same belief for all types of the same agent.

While convexity implies that giving agents no private information about the ranking is optimal, it is a priori unclear whether some ranking scheme achieves this, and moreover whether in this way the principal can achieve any ranking distribution without restriction. The proof of Theorem 1 shows that this is indeed true. For any given desired distribution, we construct a sequence of ranking schemes that approximates this distribution and gives agents no information about their rank with probability approaching one.

To understand our construction, it is instructive to specialize it to the two-worker symmetric example of Section 2. Recall that in that example, we con-

sidered an incentive scheme that would uniquely implement work by making worker 2 uncertain about his rank. Specifically, worker 2 did not know whether worker 1 was offered a high bonus b_H (making work dominant) or a low bonus b_L (making work dominant only conditional on worker 2 also working), and worker 2's bonus b_M would make him just willing to work for uniform beliefs about worker 1. Given the assumption that the project would succeed for sure under both workers working, the principal's value under this scheme was equal to $\frac{b_H+b_L}{2}+b_M$. Theorem 1 however tells us that the principal can do better; in fact, her optimal value in this simple setting is equal to $2b_M$.

We can approximate the principal's optimal value by constructing a sequence of ranking schemes that makes both workers uncertain about their rank. For each $m \in \mathbb{N}$, let the principal (privately) draw a random variable ℓ which is uniform over $\{0, \ldots, m-1\}$. Independent of ℓ and uniformly across the workers, one worker is chosen to be the "leader" and the other to be the "follower". The leader is told that his type is $\ell+1$ and the follower is told that his type is $\ell+2$; crucially, neither is told whether he is the leader or the follower. Consider a worker's ranking belief given his type in $\{1, \ldots, m+1\}$: when his type is 1, he knows he is the leader; when his type is $m+1$, he knows he is the follower; and for every other type, he assigns probability $\frac{1}{2}$ to being the leader. Therefore, the principal can uniquely implement work by specifying bonus b_H for type 1, bonus b_L for type $m+1$, and bonus b_M for every other type. Under this scheme, work is dominant for either worker who is told to be type 1 and hence, proceeding by induction on the type, for each type in $\{1, \ldots, m+1\}$.[10] Furthermore, as m becomes large, this scheme pays an average bonus arbitrarily close to b_M to each worker, thus approximating the principal's optimal value in Theorem 1.

4.3. Bonus Distribution

We have shown that the principal's optimal value can be characterized via the auxiliary optimization over ranking distributions given in Theorem 1. We now show that this auxiliary program also permits a characterization of opti-

[10] This contagion argument is reminiscent of Rubinstein's (1989) email game.

mal incentives. The following theorem fully characterizes optimal sequences of incentive schemes, by characterizing the limiting bonus distributions they generate.

Theorem 2. *There is a unique bonus profile $b^* \in \mathbb{R}^N$ which minimizes $\sum_{i \in N} b_i$ among all*

$$b \in \left\{ \tfrac{1}{P(N)} \left(f_1(\mu), \ldots, f_N(\mu) \right) : \ \mu \in \Delta\Pi \right\}.$$

Furthermore, a sequence $(\sigma^m)_m$ of incentive schemes that uniquely implement work is optimal if and only if, for every $i \in N$, the induced bonus distribution of agent i under σ^m converges to a degenerate distribution on b_i^*.

Theorem 2 shows that optimal incentive contracts are (approximately) unique, with agents' bonus payments being a continuous function of the parameters of the model.[11] It is worth noting that both of these features stand in contrast with the results of Winter (2004). In a setting in which the principal's contract offers are constrained to be public, Winter (2004) finds that the optimal incentive scheme may be arbitrary (for example, when all agents are symmetric), and so agents' payments may change discontinuously with the model parameters (for example, when agents become slightly asymmetric). Instead, allowing contract offers to be private delivers a unique and well-behaved solution.

Theorem 2 tells us that, for some parameter values, the principal benefits *strictly* from being able to contract privately with the agents. In particular, we find that as long as the agents are not too asymmetric, the principal's unique optimal incentive scheme introduces rank uncertainty. Agents not only face ex-ante uncertainty about their rank relative to others, but they also face this uncertainty interim, as they learn minimal payoff-relevant information about others from their own contract offers.

Rank uncertainty reduces the expected bonus payments that the principal has to offer to ensure work, and it also impacts the optimal distribution of

[11] Berge's theorem implies that optimal ranking distributions (as in Theorem 1) move upper hemicontinuously with the model's parameters. The uniqueness result of Theorem 2 then implies optimal bonuses are a continuous function of said parameters.

bonuses across the agents. Consider the case of symmetric agents.

Definition 2. Agents $i, j \in N$ are **symmetric** if they have the same cost of working and marginal product, i.e., $c_i = c_j$ and $P(J \cup \{i\}) = P(J \cup \{j\})$ for each $J \subseteq N \setminus \{i, j\}$.

The next corollary follows from Theorem 2:

Corollary 1. *Suppose agents $i, j \in N$ are symmetric. If a sequence $(\sigma^m)_m$ of incentives schemes that uniquely implement work is optimal, then the induced bonus distributions of agents i and j under σ^m converge to degenerate distributions on bonuses b_i^* and b_j^* with $b_i^* = b_j^*$.*

The principal's optimal incentive scheme treats symmetric agents symmetrically. Importantly, Corollary 1 says that pay discrimination is not just unnecessary but strictly suboptimal for the principal. Any two symmetric agents receive the same bonus offer and they also face the same uncertainty about their ranks. In fact, if all agents are symmetric, the principal optimally induces uniform beliefs over all possible agent rankings, so that the optimal limiting bonus is $b_i^* := \frac{1}{P(N)} f_1(\mu_!)$ for all $i \in N$ with $\mu_! \in \Delta\Pi$ uniform.

As discussed in the Introduction, these findings provide a key counterpoint to Winter's (2004) seminal discrimination result. When the principal's contract offers are public, Winter (2004) finds that any optimal incentive scheme must treat all agents differently. Such a scheme discriminates between symmetric agents in an arbitrary way (at least ex post) and it treats nearly symmetric agents disproportionally differently (both ex ante and ex post). Intuitively, as in our setting, the principal uses a ranking scheme to uniquely induce the agents to work. However, under public contracts, the agent ranking must be transparent, and it must thus be accompanied by a hierarchy of bonuses: the first agent in the ranking is offered the highest bonus so that working is dominant; the second agent is offered the second highest bonus so that working is dominant knowing that the first agent works; and so on.

Theorem 2 and Corollary 1 show that this logic breaks down when rank uncertainty is introduced. By announcing only a ranking distribution, the principal can make each agent assign positive probability to other agents working,

thereby providing each of them with some assurance that other agents will work, instead of building this assurance hierarchically as under public contracts above. We find that when agents are symmetric, the principal indeed minimizes bonus payments by building assurance symmetrically. The reason is that agents' incentives to work increase at a decreasing rate with their beliefs that other agents will work, so the principal wishes to increase the beliefs of precisely those agents who are most pessimistic. As a consequence, the principal's optimal incentive scheme makes symmetric agents hold the same belief about their rank and, in turn, require the same bonus payment to work.

4.4. Comparative Statics

[TO BE ADDED]

5. Discussion

In this section, we briefly describe some ways in which our model can be enriched, point to how the analysis changes (with formal arguments available upon request), and identify some open questions.

Strategic substitutability. While complementarities are a natural feature of team production, some substitutability might exist, either due to congestion costs or because some aspects of production may only require the efforts of a small number of workers. If such substitutability overwhelms any complementarity so that P is submodular, then our principal's problem is simple: making it an equilibrium for all agents to work necessitates making it dominant for all agents to work, because the case that all other agents are working is the worst-case belief for an agent's own marginal product. In particular, in this case, there is then no scope for beneficial rank uncertainty.

But what if, as seems more realistic when there are many agents, P is neither supermodular nor submodular? In this case, an analysis nearly identical to ours can be used to study the least-cost way to incentivize work from every agent as a unique (interim correlated) *rationalizable* choice. The main difference in the analysis is that an agent who knows that a subset J of other agents

is working may suspect that even more other agents are working, and this can make him even more pessimistic about his own marginal product. To take this possibility into account, for each $J \subseteq N$ and $i \in N \setminus J$, let

$$m_i(J) := \min \left\{ P\left(\hat{J} \cup \{i\}\right) - P(\hat{J}) : \ J \subseteq \hat{J} \subseteq N \setminus \{i\} \right\}$$

denote the minimum marginal product that agent i can entertain given that all agents $j \in J$ work.[12] With this worst-case marginal product in hand, the rest of our analysis extends readily. Specifically, redefining the cost of providing incentives to agent i (given by (2) in our analysis) via

$$f_i(\mu_i) := \frac{c_i P(N)}{\sum_{\pi \in \Pi} \mu_i(\pi) m_i\left(\{j \in N : \ \pi_j < \pi_i\}\right)},$$

all of our main results go through as stated.[13]

At this stage, it remains an open question whether the principal can more cheaply incentivize all agents to work as a unique *equilibrium* instead of a unique rationalizable choice. The reason is that different types of a given agent might have their worst-case beliefs generated by different strategies for other agents. This means that the principal may be able to implement work as a unique equilibrium at a lower cost, but also that one cannot optimize agents' bonuses separably across their types, making the analysis more delicate.

Information sharing. Our analysis, and the possibility of rank uncertainty and its benefits, relies on the principal's ability to withhold information from one agent about another agent's contract terms. A natural question, then, is how robust the principal's scheme is to agents being able to share this information. It turns out that the possibility of information sharing can completely

[12] Clearly, a minimizing $\hat{J}$ in the definition of $m_i(J)$ is simply equal to J itself (respectively, $N \setminus \{i\}$) in the case in which P is supermodular (respectively, submodular).

[13] Observe that, as claimed, the cost of incentivizing agent i in the special case in which P is submodular is

$$f_i(\mu_i) := \frac{c_i P(N)}{\sum_{\pi \in \Pi} \mu_i(\pi)[P(N) - P(N \setminus \{i\})]} = \frac{c_i P(N)}{P(N) - P(N \setminus \{i\})}$$

independent of his ranking belief.

undermine the principal's ability to leverage rank uncertainty.

To see this, let us restrict attention to the case of symmetric production (i.e., $P(J) = P\{1, \ldots, |J|\}$ for all $J \subseteq N$) and suppose that the agents have the option to verifiably disclose their type to other agents before effort choices are made. Given an incentive scheme $\sigma = \langle q, B \rangle$, we claim there is a perfect Bayesian equilibrium in which all agents disclose their types, and the lowest-effort equilibrium of the associated complete-information game (which exists and is in pure strategies) is played thereafter. For any $t \in T^q$, let $J(t)$ be the set of agents who work in said equilibrium of the induced complete-information game. Additionally, for each $i \in N$ and type profile t_{-i} of agents other than i, let $t_i^*(t_{-i}) \in T_i^q$ be chosen to minimize the effort of such agents in the complete-information game for (t_i^*, t_{-i}); that is, let $t_i^*(t_{-i}) \in \arg\min_{t_i \in T_i^q} |J(t) \setminus \{i\}|$. We can then construct an equilibrium in which, whenever an agent i fails to disclose his type, all other agents believe that agent i's type is $t_i^*(t_{-i})$, therefore punishing him for nondisclosure.

More precisely, consider the following strategy profile. Every agent discloses his type, and then exactly the agents in $J(t)$ work if type profile t was disclosed. If everybody but one agent i discloses his type, then all disclosing agents believe that agent i's type is $t_i^*(t_{-i})$, and exactly the agents in $J\left(t_i^*(t_{-i}), t_{-i}\right)$ work. If any larger group of agents fail to disclose, let agents' beliefs about the types of nondisclosing agents and resulting continuation equilibrium be arbitrary. It is straightforward that this profile constitutes an equilibrium, and that this equilibrium generates a payoff for the principal that she could have attained by using public contracts. Therefore, if agents can voluntarily, verifiably disclose their contracts to other agents, the principal cannot hope to outperform the public-contract benchmark studied by Winter (2004).

This analysis may help understand why, in reality, firms insist that employees do not share their contract terms with each other. According to survey data, a significant number of private sector employers in the US either discourage or even formally prohibit the discussion of salary information, and many workers report that discussing this information can lead to punishment; see Gely and Bierman (2003) and Hegewisch et al. (2011), among others. One

potential explanation why firms maintain these policies is that pay secrecy puts workers at an informational disadvantage when negotiating their contract terms. However, workers in many cases do not have the bargaining power to exploit information about other workers' contracts, and when they do, it is unclear that transparency would allow them to negotiate better terms (cf. Cullen and Pakzad-Hurson, 2019). We study a model in which the firm makes take-it-or-leave-it offers to the workers and show that limiting workers' information can be beneficial to the firm even in such a case. Maintaining pay secrecy benefits the firm not by improving its bargaining position, but by improving its ability to address strategic risk in incentive provision.

Interdependent contracting. In our model, an incentive scheme is required to make each agent's success-contingent bonus a function solely of his own type. This modeling choice reflects a realistic feature of employees' contract terms, namely that they are perfectly informed of the terms of their own employment. However, in settings in which bonuses are of a discretionary nature, it is conceivable that agents face uncertainty about their own pay. We study here how allowing each agent's bonus to depend on the entire profile of agent types changes the principal's problem.

Let an incentive scheme $\sigma = \langle q, B \rangle$ specify a finite-support prior $q \in \Delta(\mathbb{N}^N)$ and a bonus rule $B = (B_1, ..., B_N)$ as in our baseline model, but where the bonus for each agent $i \in N$ is now $B_i(t)$, i.e. a function not only of agent i's type $t_i \in T_i^q$ but of the entire type profile $t \in T^q$. We begin by noting that this distinction is immaterial if the principal only wants to implement every agent working as *some* equilibrium. Such an equilibrium requires that, conditional on all other agent types working, each agent prefer to work whatever his type than to shirk whatever his type: his average success-contingent bonus must be at least

$$\sum_{t \in T^q} q(t) B_i(t) \geq \frac{c_i}{P(N) - P(N \setminus \{i\})} =: b_i^L.$$

Clearly, publicly offering each agent a success-contingent bonus of b_i^L yields every agent working as an equilibrium, so the possibility of making agents

uncertain about other agents' or their own contract terms does not matter for this design problem.

But what if, as motivates our study, the principal wishes to implement every agent working as a *unique* equilibrium? Of course, she can do no better than having to pay only $P(N)b_i^L$ on average to each agent $i \in N$. We next show that when agents' bonuses can depend on the entire type profile, the principal can, in fact, approximate this theoretical bound arbitrarily well.

For given $\gamma \in (0,1)$, consider the following incentive scheme. Each agent $i \in N$ is assigned type $t_i = 1$ with probability γ and type $t_i = 2$ with probability $1-\gamma$, independent across agents. The type *profile*-contingent bonus rule is given by

$$B_i(t) := \begin{cases} \frac{c_i}{P\{i\}-P(\emptyset)} & : \ t_i = 1, \\ \frac{1}{\gamma^{N-1}} \cdot b_i^L & : \ t_i = 2 \text{ and } t_j = 1 \text{ for all } j \in N \setminus \{i\}, \\ 0 & : \ \text{otherwise.} \end{cases}$$

Observe that this scheme uniquely implements work: a type-1 agent finds it dominant to work, and a type-2 agent wants to work given that all other agents work when assigned type 1 (since the probability that all others are type 1 is γ^{N-1}). Furthermore, the expected payment to each agent $i \in N$ is

$$P(N)\left[\gamma \cdot \frac{c_i}{P\{i\}-P(\emptyset)} + (1-\gamma)\gamma^{N-1} \cdot \frac{1}{\gamma^{N-1}} b_i^L\right],$$

which converges to $P(N)b_i^L$ as $\gamma \to 0$.

We thus find that if the principal is not required to inform agents of their own contractual terms, then she can implement work while excluding other equilibria at no extra cost. Allowing for contracts to be private in this (possibly unrealistically) stronger sense makes the unique implementation requirement essentially lose all of its bite.[14]

[14] The idea that complex enough contracts can allow one to exclude undesirable equilibria at no additional cost arises in other models of contracting with externalities. In particular, in settings in which agents' choices are contractible, this is true if contracts are not required to be bilateral but can be made contingent on third parties' choices; see Segal (1999, 2003).

References

Bernstein, Shai and Eyal Winter, "Contracting with Heterogeneous Externalities," *American Economic Journal: Microeconomics*, 2012, *4*, 50–76.

Cullen, Zoë B. and Bobak Pakzad-Hurson, "Equilibrium Effects of Pay Transparency in a Simple Labor Market," 2019. Working Paper.

_ **and Ricardo Perez-Truglia**, "The Salary Taboo: Privacy Norms and the Diffusion of Information," 2018. NBER Working Paper No. 25145.

Doval, Laura and Jeff Ely, "Sequential Information Design," 2019. Working Paper.

Edwards, Matthew A., "The Law and Social Norms of Pay Secrecy," *Berkeley Journal of Employment and Labor Law*, 2005, *26* (1), 41–63.

Eliaz, Kfir and Ran Spiegler, "X-games," *Games and Economic Behavior*, 2015, *89*, 93–100.

Gely, Rafael and Leonard Bierman, "Pay Secrecy/Confidentiality Rules and the National Labor Relations Act," *Journal of Labor and Employment Law*, 2003, *6*, 120–156.

Habibi, Amir, "Pay Transparency in Organisations," 2019. Working Paper.

Halac, Marina, Ilan Kremer, and Eyal Winter, "Raising Capital from Heterogeneous Investors," *American Economic Review*, forthcoming.

Hegewisch, Ariane, Claudia Williams, and Robert Drago, "Pay Secrecy and Wage Discrimination," *Institute for Women's Policy Research*, 2011.

Holmström, Bengt, "Moral Hazard in Teams," *The Bell Journal of Economics*, 1982, *13* (2), 324–340.

Hoshino, Tetsuya, "Multi-Agent Persuasion: Leveraging Strategic Uncertainty," 2019. Working Paper.

Inostroza, Nicolas and Alessandro Pavan, "Persuasion in Global Games with Application to Stress Testing," 2018. Working Paper.

Mathevet, Laurent, Jacopo Perego, and Ina Taneva, "On Information Design in Games," *Journal of Political Economy*, forthcoming.

Milgrom, Paul and John Roberts, "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 1990, pp. 1255–1277.

Moriya, Fumitoshi and Takuro Yamashita, "Asymmetric-Information Allocation to Avoid Coordination Failure," *Journal of Economics & Management Strategy*, 2019, *forthcoming*.

Morris, Stephen, Daisuke Oyama, and Satoru Takahashi, "Adversarial Information Design in Binary-Action Supermodular Games," 2019. Working Paper.

Rubinstein, Ariel, "The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge"," *American Economic Review*, 1989, *79* (3), 385–391.

Segal, Ilya R., "Contracting with Externalities," *Quarterly Journal of Economics*, 1999, *114*, 337–388.

_ , "Coordination and Discrimination in Contracting with Externalities: Divide and Conquer?," *Journal of Economic Theory*, 2003, *113*, 147–81.

Shellenbarger, Sue, "Open Salaries: the Good, the Bad and the Awkward," *The Wall Street Journal*, 2016, *January 12*.

Trotter, Richard G., Susan Rawson Zacur, and Lisa T. Stickney, "The New Age of Pay Transparency," *Business Horizons*, 2017, *60* (4), 529–539.

Winter, Eyal, "Incentives and Discrimination," *American Economic Review*, 2004, *94*, 764–773.

_ , "Optimal Incentives for Sequential Production Processes," *The RAND Journal of Economics*, 2006, *37* (2), 376–390.

_ , "Transparency and Incentives Among Peers," *The RAND Journal of Economics*, 2010, *41* (3), 504–523.

A. Omitted Proofs

A.1. Proof of Proposition 1

We begin by proving the first part, namely that every ranking scheme uniquely implements work. Fix a ranking scheme q and an $\varepsilon > 0$. Consider an arbitrary Bayesian Nash equilibrium of the Bayesian game induced by $\langle q, B + \varepsilon \rangle$. By definition of a ranking scheme, every type t_i of every agent $i \in N$ has work as his unique best response given bonus $\frac{1}{P(N)} f_i(t_i) + \varepsilon$ if exactly all strictly lower types of other agents do, and so, by supermodularity of P, if at least all strictly lower types do. Therefore, it follows by induction on the type that each type t_i of every agent $i \in N$ works in the equilibrium.

We next prove the second part of the proposition. Fix an incentive scheme $\sigma = \langle q, B \rangle$ that uniquely implements work. We proceed in three steps.

Step 1. We assign each agent type a finite index. Define the profile of functions $k_i : T_i^q \to \mathbb{N} \cup \{\infty\}$ for $i \in N$ as follows. Let $T_i^q(0) := \emptyset$, and then, recursively for $\kappa \in \mathbb{N}$, let $T_i^q(\kappa)$ be the set of all $t_i \in T_i^q$ such that

$$B_i(t_i) \sum_{t_{-i}} q_i(t_{-i}|t_i) \left[\begin{array}{c} P\left(\{j \in N : \ t_j \in T_j^q(\kappa - 1)\} \cup \{i\}\right) \\ -P\left(\{j \in N : \ t_j \in T_j^q(\kappa - 1)\} \setminus \{i\}\right) \end{array} \right] \geq c_i.$$

Define $k_i(t_i) := \inf\{\kappa \in \mathbb{N} : \ t_i \in T_i^q(\kappa)\}$. We show that all indices $\{k_i\}_{i \in N}$ are finite-valued. As T^q is finite, there is some $\overline{\kappa} \in \mathbb{N}$ strictly higher than everything in $\bigcup_{i \in N} k_i(T_i^q) \setminus \{\infty\}$. Now, for each agent $i \in N$ and each type $t_i \in k_i^{-1}(\infty)$, that $k_i(t_i) > \overline{\kappa}$ implies

$$\varepsilon_i(t_i) := c_i - B_i(t_i) \sum_{t_{-i}} q_i(t_{-i}|t_i) \left[\begin{array}{c} P\left(\{j \in N : \ k_j(t_j) < \infty\} \cup \{i\}\right) \\ -P\left(\{j \in N : \ k_j(t_j) < \infty\} \setminus \{i\}\right) \end{array} \right] > 0.$$

Letting $\varepsilon := \min\{\varepsilon_i(t_i) : i \in N, t_i \in k_i^{-1}(\infty)\} > 0$, observe that the Bayesian game induced by prior q and bonuses $B_i + \varepsilon$ has a Bayesian Nash equilibrium in which each agent i of type t_i works if $k_i(t_i) < \infty$ and shirks if $k_i(t_i) = \infty$. Indeed, the best response property for type t_i follows directly from the definition of $k_i(t_i)$ when $k_i(t_i) < \infty$, and it follows additionally from the definition of ε when $k_i(t_i) = \infty$. But then, that σ uniquely implements work implies $k_i^{-1}(\infty) = \emptyset$ for every agent $i \in N$, as desired.

Step 2. Using the indices from Step 1, we construct a ranking scheme $\sigma^* = \langle q^*, B^* \rangle$ from the original scheme $\sigma = \langle q, B \rangle$. First, note that we can construct a one-to-one function

$$\lambda : \bigcup_{i \in N} [\{i\} \times T_i^q] \to \mathbb{N}$$

such that, for any $i, j \in N$ and $t_i \in T_i^q$, $t_j \in T_j^\sigma$ with $k_i(t_i) > k_i(t_j)$, we have $\lambda(i, t_i) > \lambda(j, t_j)$.[15] Then, using this function, construct the incentive scheme $\sigma^* = \langle q^*, B^* \rangle$ as follows. Let the prior q^* be given by

$$q^*(t^*) = \begin{cases} q(t) & : t \in T^q \text{ has } \lambda(i, t_i) = t_i^* \text{ for all } i \in N, \\ 0 & : \text{otherwise.} \end{cases}$$

Let the bonus rule $B^* = (B_i^*)_{i \in N}$ be given by $B_i^*(t_i^*) := \frac{1}{P(N)} f_i\left(\mu_i^{q^*}(t_i^*)\right)$. Injectivity of λ ensures that, with q^*-distributed type profiles, no two agents ever have the same type realized. It follows directly that q^* is a ranking scheme.

Step 3. All that remains is to show that the bonus distributions induced by the ranking scheme q^* constructed in Step 2 are first-order-stochastically dominated by those induced by the original scheme $\sigma = \langle q, B \rangle$. As q^* is constructed from q by relabeling types and leaving their respective probabilities unchanged, it suffices to check that every $i \in N$ and $t_i \in T_i^q$ has $B_i(t_i) \geq$

[15] For instance, letting $L := 1 + n + \max_{i \in N, t_i \in T_i^\sigma} t_i$, take $\lambda(i, t_i) := L^2 k_i(t_i) + Li + t_i$.

$B_i^*(\lambda(i,t_i))$. Letting $t_i^* := \lambda(i,t_i)$, we indeed verify that

$$\begin{aligned} B_i(t_i) &\geq \frac{c_i}{\sum_{t_{-i}} q_i(t_{-i}|t_i)[P(\{j\in N\colon k_j(t_j)<k_i(t_i)\}\cup\{i\})-P\{j\in N\colon k_j(t_j)<k_i(t_i)\}]} \\ &\geq \frac{c_i}{\sum_{t_{-i}} q_i(t_{-i}|t_i)[P(\{j\in N\colon \lambda(j,t_j)<\lambda(i,t_i)\}\cup\{i\})-P\{j\in N\colon \lambda(j,t_j)<\lambda(i,t_i)\}]} \\ &= \frac{c_i}{\sum_{t^*_{-i}} q_i^*(t^*_{-i}|t_i^*)\left[P\left(\{j\in N\colon t_j^*<t_i^*\}\cup\{i\}\right)-P\{j\in N\colon t_j^*<t_i^*\}\right]} \\ &= B_i^*(t_i^*), \end{aligned}$$

where the first inequality follows from the definition of $k_i(t_i)$, and the second inequality follows from supermodularity of P and the defining feature of λ.

A.2. Proof of Theorem 2

Before proving the theorem, we record the following lemma about the function $f_i : \Delta\Pi \to \mathbb{R}_{++}$ that describes the expected fees required to incentivize agent $i \in N$.

Lemma 1. *For each $i \in N$, every $\tau_i \in \Delta\Delta\Pi$ has $\int f_i \, d\tau_i \geq f_i\left(\int \mu_i \, d\tau_i(\mu_i)\right)$, with a strict inequality if τ_i induces a nondegenerate distribution of f_i.*

Proof. The function $\mathbb{R}_{++} \to \mathbb{R}$ given by $x \mapsto \frac{1}{x}$ is strictly convex and continuous, while the function $\frac{1}{f_i}$ is affine and continuous. Therefore, Jensen's inequality yields

$$\int f_i \, \mathrm{d}\tau_i = \int \frac{1}{\left(\frac{1}{f_i}\right)} \, \mathrm{d}\tau_i \geq \frac{1}{\int \frac{1}{f_i} \, \mathrm{d}\tau_i} = \frac{1}{\frac{1}{f_i}\left(\int \mu_i \, \mathrm{d}\tau_i(\mu_i)\right)} = f_i\left(\int \mu_i \, \mathrm{d}\tau_i(\mu_i)\right),$$

where the inequality is strict if the τ_i-distribution of f_i (and thereby $\frac{1}{f_i}$) is nondegenerate. *Q.E.D.*

We now proceed to prove the theorem. Consider the first claim, which says that there is a unique μ^* which minimizes $\sum_{i\in N} f_i(\mu)$ among all $\mu \in \Delta\Pi$. Assume for a contradiction that this is not true. Since $\sum_{i\in N} f_i(\mu)$ attains a minimum somewhere, there must then be two minimizers $\mu', \mu'' \in \Delta\Pi$ of $\sum_i f_i(\mu)$ with $(f_i(\mu'))_{i\in N} \neq (f_i(\mu''))_{i\in N}$. Applying Lemma 1 to a uniform mixture over

$\{\mu', \mu''\}$ yields

$$\sum_{i \in N} f_i\left(\tfrac{1}{2}\mu' + \tfrac{1}{2}\mu''\right) < \sum_{i \in N}\left(\tfrac{1}{2}f_i(\mu') + \tfrac{1}{2}f_i(\mu'')\right) = \tfrac{1}{2}\sum_{i \in N} f_i(\mu') + \tfrac{1}{2}\sum_{i \in N} f_i(\mu''),$$

contradicting that $\mu', \mu'' \in \Delta\Pi$ are minimizers. Hence, there is a unique minimizer as claimed.

Consider now the second claim. The "if" direction of this claim follows immediately from Theorem 1. To prove the "only if" direction, consider an optimal sequence $(\sigma^m)_m$ of incentive schemes. Apply Proposition 1 to obtain, for each $m \in \mathbb{N}$, a ranking scheme $\hat{q}^m$ such that $\beta_i^{\sigma^m}$ first-order-stochastically dominates the bonus distribution $\beta_i^{\hat{q}^m}$ for each $i \in N$. Notice that $\beta_i^{\hat{q}^m}$ is supported on the compact set $\left[0, \frac{c_i}{P\{i\}-P(\emptyset)}\right]$ by the definition of a ranking scheme. We therefore need only prove that the unique limit point of the sequence $(\beta_i^{\hat{q}^m})_m$ is degenerate on b_i^* for all $i \in N$. Indeed, because of first-order stochastic dominance and optimality of $(\sigma^m)_m$, it follows that $(\beta_i^{\sigma^m})_m$ converges to the same degenerate bonus distribution.

For each $i \in N$ and each $m \in \mathbb{N}$, the ranking scheme $\hat{q}^m$ generates a finite-support distribution $\tau_i^m \in \Delta\Delta\Pi$ over agent i's ranking belief, given by

$$\tau_i^m(\mu_i) := \hat{q}_i^m\{t_i : \ \mu_i^{\hat{q}^m}(t_i) = \mu_i\} \text{ for all } \mu_i \in \Delta\Pi.$$

Note that any limit point β_i of $(\beta_i^{\hat{q}^m})_m$ exhibits some limit point of $(\tau_i^m)_m$ whose induced distribution of f_i is β_i. We therefore focus on an arbitrary limit point of $(\tau_i^m)_m$, denoted τ_i^*. Since each agent has Bayes-consistent beliefs about the ranking distribution,

$$\int \mu_i \, \mathrm{d}\tau_i^m(\mu_i) = \sum_{t_i} \hat{q}_i^m(t_i)\mu_i^{\hat{q}^m}(t_i) = \mu^{\hat{q}^m}$$

is independent of i. Therefore, taking limits, $\int \mu_i \, \mathrm{d}\tau_i^*(\mu_i) =: \mu^*$ is also independent of i. We next observe that τ_i^* must induce a degenerate distribution

of f_i for every $i \in N$. Indeed, if not, then Lemma 1 would imply

$$\lim_{m\to\infty} W(\hat{q}^m) = \sum_{i\in N} \int f_i \, \mathrm{d}\tau_i^* > \sum_{i\in N} f_i(\mu^*) \geq \min_{\mu\in\Delta\Pi} \sum_{i\in N} f_i(\mu),$$

which (given Theorem 1) would violate the optimality of $(\hat{q}^m)_m$.

Hence, we obtain that $(\tau_i^m)_m$ converges to a degenerate distribution on μ^* for every agent $i \in N$, implying that $\lim_{m\to\infty} W(\hat{q}^m) = \sum_{i\in N} f_i(\mu^*)$. By the first claim in this theorem and the result in Theorem 1, optimality then requires $f_i(\mu^*) = P(N)b_i^*$ for all $i \in N$. This completes the proof.